PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-101226
(43)Date of publication of application : 13.04.2001

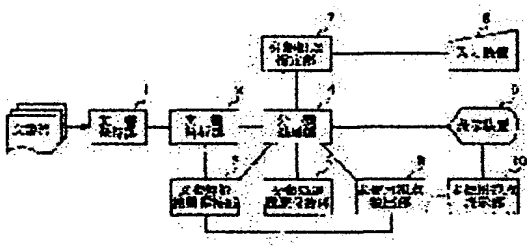
(51)Int.Cl. G06F 17/30

(21)Application number : 11-282013
(22)Date of filing : 01.10.1999

(71)Applicant : RICOH CO LTD
(72)Inventor : YAMAZAKI MAKOTO

(54) DOCUMENT GROUP SORTER AND DOCUMENT GROUP SORTING METHOD

(57)Abstract:
PROBLEM TO BE SOLVED: To provide document group sorter, etc., capable of obtaing a desired document subset by confirming a document sorting viewpoint which is used so far and the unused one.
SOLUTION: This document group sorter to sort a document set according to the contents of documents is provided with a document analyzing part 2 to extract information required for a sorting processing by analyzing document data of each document set as a sorting object, a sorting processing specifying part 7 to specify a sorting viewpoint, etc., in the case of the sorting processing, a sorting processing part 4 to sort the document set into plural document subsets according to the information extracted by the document analyzing part 2 and the sorting viewpoint specified by the sorting processing specifying part 7, a sorting processing history holding part 5 to hold history information of a sorting processing result, an unused viewpoint detecting part 8 to detect the sorting viewpoint which is included in the document set but still unused based on the history information and an unused viewpoint display part 10 to display the detected sorting viewpoint when the sorting viewpoint is specified.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報（Ａ）**　(11)特許出願公開番号
特開2001－101226
（P2001－101226A）
(43)公開日　平成13年４月13日(2001.4.13)

(51)Int.Cl.[7]	識別記号	ＦＩ		テーマコード（参考）
Ｇ０６Ｆ　17/30		Ｇ０６Ｆ　15/401	３１０Ｄ	５Ｂ０７５
		15/40	３７０Ａ	
		15/403	３５０Ｃ	

審査請求　未請求　請求項の数16　ＯＬ　（全　９　頁）

(21)出願番号　特願平11－282013

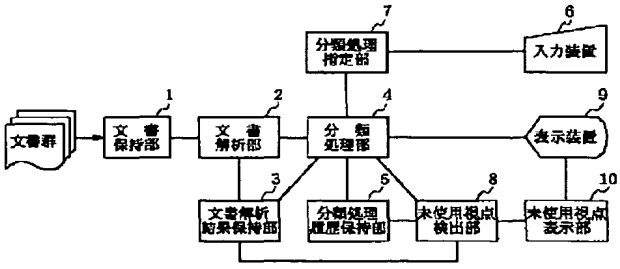
(22)出願日　平成11年10月１日(1999.10.1)

(71)出願人　000006747
株式会社リコー
東京都大田区中馬込１丁目３番６号
(72)発明者　山崎　真湖人
東京都大田区中馬込１丁目３番６号　株式会社リコー内

Ｆターム(参考)　5B075 NR02 NR12 PP02 PP03 PQ02 PR04 PR06 QM08 UU06

(54)【発明の名称】　文書群分類装置および文書群分類方法

(57)【要約】
【課題】　これまでに用いられた文書分類視点やまだ用いられていない文書分類視点の確認を可能にして所望の文書部分集合を得られるようにした文書群分類装置などを提供する。
【解決手段】　文書の内容に従って文書集合を分類する文書群分類装置において、分類対象の文書集合のそれぞれの文書データを解析して分類処理に必要な情報を抽出する文書解析部２、分類処理に際して分類視点などを指定させる分類処理指定部７、文書解析部２により抽出された情報および分類処理指定部７により指定された分類視点に従って文書集合を複数の文書部分集合に分類する分類処理部４、分類処理結果の履歴情報を保持する分類処理履歴保持部５、前記履歴情報に基づいて前記文書集合に内在するがまだ用いていない分類視点を検出する未使用視点検出部８、検出された前記分類視点を分類視点指定時に表示させる未使用視点表示部10を備えた。



【特許請求の範囲】

【請求項１】　文書の内容に従って文書集合を分類する文書群分類装置において、複数の文書から成る文書集合のそれぞれの文書データを保持する文書保持手段と、前記文書保持手段に保持されたそれぞれの文書データを解析して分類処理に必要な情報を抽出する文書解析手段と、分類処理に際して分類視点を指定する分類視点指定手段と、前記文書解析手段により抽出された情報および前記分類視点指定手段により指定された分類視点に従って文書集合を複数の文書部分集合に分類する分類処理手段と、前記分類処理手段による分類処理結果の履歴情報を保持する分類処理履歴保持手段と、前記分類処理履歴保持手段に保持された分類処理結果の履歴情報に基づいて前記文書集合に内在するがまだ用いていない分類視点を検出する未使用視点検出手段と、前記分類視点指定手段による分類視点指定時に前記未使用視点検出手段により検出された用いられていない分類視点を表示させる未使用視点表示手段とを備えたことを特徴とする文書群分類装置。

【請求項２】　請求項１記載の文書群分類装置において、指定された分類視点情報およびその分類視点に従った分類結果である文書部分集合情報を保持するように分類処理履歴保持手段を構成したことを特徴とする文書群分類装置。

【請求項３】　請求項１または請求項２記載の文書群分類装置において、未使用視点表示手段により表示された分類視点情報を用いて分類視点を指定させるように分類視点指定手段を構成したことを特徴とする文書群分類装置。

【請求項４】　請求項３記載の文書群分類装置において、さらに、分類視点情報を含んだ分類処理結果履歴情報を表示させ、表示された履歴情報中の分類視点を用いて分類視点を指定させるように分類視点指定手段を構成したことを特徴とする文書群分類装置。

【請求項５】　請求項１乃至請求項４記載の文書群分類装置において、未使用視点情報を表示させる際、それぞれの未使用視点を示す一つ以上の単語および／または前記未使用視点の文書部分集合を代表する文書を示す情報を表示させるように未使用視点表示手段を構成したことを特徴とする文書群分類装置。

【請求項６】　請求項１乃至請求項５記載の文書群分類装置において、記憶しておいた分類視点情報を含む情報を表示させ、表示された分類視点情報を編集する分類視点編集手段を備え、編集された分類視点に従って分類処理を行うように分類処理手段を構成したことを特徴とする文書群分類装置。

【請求項７】　請求項６記載の文書群分類装置において、編集する分類視点情報を含む情報を未使用視点情報または分類処理結果履歴情報とする構成にしたことを特徴とする文書群分類装置。

【請求項８】　文書の内容に従って文書集合を分類する文書群分類方法において、複数の文書から成る文書集合のそれぞれの文書データを保持し、前記それぞれの文書データを解析して分類処理に必要な内在情報を抽出しておき、分類処理に際して分類視点を指定し、前記内在情報および指定された前記分類視点に従って文書集合を複数の文書部分集合に分類し、分類処理結果の履歴情報を保持しておき、保持された前記履歴情報に基づいて前記文書集合に内在するがまだ用いていない分類視点を検出するようにして、前記分類視点指定時に、前記検出された用いられていない分類視点を表示させることを特徴とする文書群分類方法。

【請求項９】　請求項８記載の文書群分類方法において、分類処理結果の履歴情報として、指定された分類視点情報およびその分類視点に従った分類結果である文書部分集合情報を保持することを特徴とする文書群分類方法。

【請求項１０】　請求項８または請求項９記載の文書群分類方法において、表示された用いられていない分類視点情報を用いて分類視点を指定させることを特徴とする文書群分類方法。

【請求項１１】　請求項10記載の文書群分類方法において、さらに、分類視点情報を含んだ分類処理結果履歴情報を表示させ、表示された履歴情報中の分類視点を用いて分類視点を指定させることを特徴とする文書群分類方法。

【請求項１２】　請求項８乃至請求項11記載の文書群分類方法において、未使用視点情報を表示させる際、それぞれの未使用視点を示す一つ以上の単語および／または前記未使用視点の文書部分集合を代表する文書を示す情報を表示させることを特徴とする文書群分類方法。

【請求項１３】　請求項８乃至請求項12記載の文書群分類方法において、記憶しておいた分類視点情報を含む情報を表示させ、表示された分類視点情報を編集させ、編集された分類視点に従って分類処理を行うことを特徴とする文書群分類方法。

【請求項１４】　請求項13記載の文書群分類方法において、編集する分類視点情報を含む情報を未使用視点情報または分類処理結果履歴情報としたことを特徴とする文書群分類方法。

【請求項１５】　請求項８乃至請求項14記載の文書群分類方法において、分類視点を引き出した文書集合と前記分類視点に従って文書分類を行う文書集合とが、異なる文書集合であることを特徴とする文書群分類方法。

【請求項１６】　プログラムを記憶した記憶媒体において、請求項８乃至請求項15記載の文書群分類方法に従ってプログラミングしたプログラムを記憶する構成にしたことを特徴とする記憶媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、文書群を文書の内容に従って複数の文書部分集合に分類する文書群分類装置などに係わり、特に、これまでに用いられた文書分類視点やまだ用いられていない文書分類視点の確認を可能にして所望の文書部分集合を得られるようにした文書群分類装置などに関する。

【０００２】

【従来の技術】近年、インターネットなどの普及により、大量の文書群へのアクセスが可能になり、その結果、そのような文書群を様々な利用者の意図に基づいて、且つ効率的に利用できるようにする必要性が高まっている。そのため、大量の文書群を意味のある文書部分集合（クラスタ）に分類するというような作業が行われ始めている。しかし、このような分類作業を人的に行おうとすると、その人的／時間的コストが膨大なものになるとか、また、分類のための知識を分類者のみが有することになるため、分類担当者が代わると分類基準も変わってしまうというような問題がある。そのため、文書群を人間が分類するような分類基準を用いて自動的に分類しうる文書分類装置が望まれるようになり、特開平7-114572号公報に示されているように、文書に含まれるそれぞれの単語の出現頻度から特徴ベクトルを抽出して、文書群を複数の文書部分集合（クラスタ）に分類する技術などが提供されるに至っている。しかし、それぞれの文書には多元的な情報が含まれているので、前記のような自動分類では利用者の意図した文書部分集合が得られないというような場合がある。そのため、分類の際に、利用者が分類視点を指定するというような方法も提供されるに至った。例えば分類視点として一つまたは複数の特定の単語を指定して指定した単語を含む文書（あるいは指定した単語を多く含む文書）を文書部分集合として分類（抽出）するのである。なお、特開平11-15835号公報に示された分類情報提示装置では、刻々と変化していく情報群に対して行われた分類の履歴を保持して表示することにより、情報群の分布がどのように変化しているかという推移情報を把握できるようにしている。

【０００３】

【発明が解決しようとする課題】しかしながら、文書分類視点を指定できるようにした前記の従来技術や、特開平11-15835号公報に示された従来技術においては、そのときまでの分類処理において用いられた文書分類視点や、文書集合に内在するがまだ用いていない文書分類視点を利用者が確認することができないので、文書分類視点の指定が一面的になってしまい、必ずしも所望の文書部分集合を得られないという問題がある。本発明の課題は、このような従来技術の問題を解決し、これまでに用いられた文書分類視点やまだ用いられていない文書分類視点の確認を可能にして所望の文書部分集合を得られるようにした文書群分類装置を提供することにある。

【０００４】

【課題を解決するための手段】前記の課題を解決するために、請求項１記載の発明では、文書の内容に従って文書集合を分類する文書群分類装置において、複数の文書から成る文書集合のそれぞれの文書データを保持する文書保持手段と、前記文書保持手段に保持されたそれぞれの文書データを解析して分類処理に必要な情報を抽出する文書解析手段と、分類処理に際して分類視点を指定する分類視点指定手段と、前記文書解析手段により抽出された情報および前記分類視点指定手段により指定された分類視点に従って文書集合を複数の文書部分集合に分類する分類処理手段と、前記分類処理手段による分類処理結果の履歴情報を保持する分類処理履歴保持手段と、前記分類処理履歴保持手段に保持された分類処理結果の履歴情報に基づいて前記文書集合に内在するがまだ用いていない分類視点を検出する未使用視点検出手段と、前記分類視点指定手段による分類視点指定時に前記未使用視点検出手段により検出された用いられていない分類視点を表示させる未使用視点表示手段とを備えた。また、請求項２記載の発明では、請求項１記載の発明において、指定された分類視点情報およびその分類視点に従った分類結果である文書部分集合情報を保持するように分類処理履歴保持手段を構成した。また、請求項３記載の発明では、請求項１または請求項２記載の発明において、未使用視点表示手段により表示された分類視点情報を用いて分類視点を指定させるように分類視点指定手段を構成した。また、請求項４記載の発明では、請求項３記載の発明において、さらに、分類視点情報を含んだ分類処理結果履歴情報を表示させ、表示された履歴情報中の分類視点を用いて分類視点を指定させるように分類視点指定手段を構成した。また、請求項５記載の発明では、請求項１乃至請求項４記載の発明において、未使用視点情報を表示させる際、それぞれの未使用視点を示す一つ以上の単語および／または前記未使用視点の文書部分集合を代表する文書を示す情報を表示させるように未使用視点表示手段を構成した。また、請求項６記載の発明では、請求項１乃至請求項５記載の発明において、記憶しておいた分類視点情報を含む情報を表示させ、表示された分類視点情報を編集する分類視点編集手段を備え、編集された分類視点に従って分類処理を行うように分類処理手段を構成した。また、請求項７記載の発明では、請求項６記載の発明において、編集する分類視点情報を含む情報を未使用視点情報または分類処理結果履歴情報とする構成にした。

【０００５】また、請求項８記載の発明では、文書の内容に従って文書集合を分類する文書群分類方法において、複数の文書から成る文書集合のそれぞれの文書データを保持し、前記それぞれの文書データを解析して分類処理に必要な内在情報を抽出しておき、分類処理に際して分類視点を指定し、前記内在情報および指定された前記分類視点に従って文書集合を複数の文書部分集合に分

類し、分類処理結果の履歴情報を保持しておき、保持された前記履歴情報に基づいて前記文書集合に内在するがまだ用いていない分類視点を検出するようにして、前記分類視点指定時に、前記検出された用いられていない分類視点を表示させる方法にした。また、請求項９記載の発明では、請求項８記載の発明において、分類処理結果の履歴情報として、指定された分類視点情報およびその分類視点に従った分類結果である文書部分集合情報を保持する方法にした。また、請求項10記載の発明では、請求項８または請求項９記載の発明において、表示された用いられていない分類視点情報を用いて分類視点を指定させる方法にした。また、請求項11記載の発明では、請求項10記載の発明において、さらに、分類視点情報を含んだ分類処理結果履歴情報を表示させ、表示された履歴情報中の分類視点を用いて分類視点を指定させる方法にした。また、請求項12記載の発明では、請求項８乃至請求項11記載の発明において、未使用視点情報を表示させる際、それぞれの未使用視点を示す一つ以上の単語および／または前記未使用視点の文書部分集合を代表する文書を示す情報を表示させる方法にした。また、請求項13記載の発明では、請求項８乃至請求項12記載の発明において、記憶しておいた分類視点情報を含む情報を表示させ、表示された分類視点情報を編集させ、編集された分類視点に従って分類処理を行う方法にした。また、請求項14記載の発明では、請求項13記載の発明において、編集する分類視点情報を含む情報を未使用視点情報または分類処理結果履歴情報とした。また、請求項15記載の発明では、請求項８乃至請求項14記載の発明において、分類視点を引き出した文書集合と前記分類視点に従って文書分類を行う文書集合とが、異なる文書集合である方法にした。また、請求項16記載の発明では、プログラムを記憶した記憶媒体において、請求項８乃至請求項15記載の文書群分類方法に従ってプログラミングしたプログラムを記憶する構成にした。

【０００６】前記のような手段にしたので、請求項１および請求項８記載の発明では、文書集合のそれぞれの文書データが解析されて分類処理に必要な内在情報が抽出しておかれ、分類処理に際して分類視点を指定すると、前記内在情報および指定された前記分類視点に従って文書集合が複数の文書部分集合に分類され、分類処理結果の履歴情報が保持され、保持された前記履歴情報に基づいて前記文書集合に内在するがまだ用いていない分類視点が検出され、その結果、前記分類視点指定時に、検出された用いられていない分類視点を表示させるようにすることができる。請求項２および請求項９記載の発明では、請求項１または請求項８記載の発明において、分類処理結果の履歴情報として、指定された分類視点情報およびその分類視点に従った分類結果である文書部分集合情報が保持される。請求項３および請求項10記載の発明では、請求項１または請求項２、または請求項８または請求項９記載の発明において、表示された用いられていない分類視点情報を用いて分類視点が指定される。請求項４および請求項11記載の発明では、請求項３または請求項10記載の発明において、さらに、分類視点情報を含んだ分類処理結果履歴情報が表示され、表示された履歴情報中の分類視点を用いて分類視点が指定される。請求項５および請求項12記載の発明では、請求項１乃至請求項４または請求項８乃至請求項11記載の発明において、未使用視点情報が表示される際、それぞれの未使用視点を示す一つ以上の単語および／または前記未使用視点の文書部分集合を代表する文書を示す情報が表示される。請求項６および請求項13記載の発明では、請求項１乃至請求項５または請求項８乃至請求項12記載の発明において、記憶しておいた分類視点情報を含む情報が表示され、表示された分類視点情報が編集され、編集された分類視点に従って分類処理が行われる。請求項７および請求項14記載の発明では、請求項６または請求項13記載の発明において、未使用視点情報または分類処理結果履歴情報が表示され、編集される。請求項15記載の発明では、請求項８乃至請求項14記載の発明において、分類視点を引き出した文書集合とは異なった文書集合に対して前記分類視点に従った文書分類が行われる。請求項16記載の発明では、請求項８乃至請求項15記載の文書群分類方法に従ってプログラミングしたプログラムが例えば着脱可能な記憶媒体に記憶される。

【０００７】

【発明の実施の形態】以下、図面により本発明の実施の形態を詳細に説明する。図１は本発明の一実施形態を示す文書分類装置の構成ブロック図である。図示したように、この実施形態の文書分類装置は、複数の文書から成る文書集合（文書群）のそれぞれの文書データを保持する文書保持手段である文書保持部１、前記文書保持部１に保持されたそれぞれの文書データを解析して分類処理に必要な内在情報を抽出する文書解析手段である文書解析部２、前記文書解析部２による解析結果情報（内在情報）を保持する文書解析結果保持部３、前記文書解析部２により抽出された情報に従って文書集合を複数の文書部分集合に分類する分類処理手段である分類処理部４、前記分類処理部４による分類処理結果の履歴情報を保持する分類処理履歴保持手段である分類処理履歴保持部５、キーボードやマウスなどから成る入力装置６、前記入力装置６と共に前記分類処理に際して分類視点を指定する分類視点指定手段などを構成する分類処理指定部７、前記分類履歴保持部５に保持された分類処理結果の履歴情報に基づいて前記文書集合に内在するがまだ用いていない分類視点を検出する未使用視点検出手段である未使用視点検出部８、表示装置９、前記分類視点指定手段による分類視点指定時に前記未使用視点検出部８により検出された、まだ用いられていない分類視点などを表示装置８に表示させる未使用視点表示手段である未使用

視点表示部10などを備えている。なお、前記文書解析部２、分類処理部４、分類処理指定部７、未使用視点検出部８、未使用視点表示部10はプログラムを記憶したメモリおよびそのプログラムに従って動作するＣＰＵを共有する。また、前記文書保持部１、文書解析結果保持部３、分類処理履歴保持部５は例えばハードディスク装置およびＲＡＭの部分領域として実現される。以下、前記各部についてさらに説明する。まず、分類対象となる複数の文書（文書集合）の文書データが保持される文書保持部１であるが、この保持方式としては、文書データベース方式やリレーショナルデータベース方式などを用いる。なお、文書データベース方式とは、例えば各文書のインデックス情報として、文書番号、文書名、登録日、登録者名、キーワードなどを登録し、そのインデックス情報に対応付けて文書コンテンツを保管するようにした方式である。

【０００８】文書解析部２は、それぞれの文書データ中から単語（例えば名詞）を抽出し、例えば個々の単語を軸とした特徴空間に表現されるそれぞれの文書に対応した特徴ベクトルを求める。つまり、文書解析部２が個々の文書データに対して言語処理を行って文書データを単語に分け、それぞれの単語の出現頻度を計数し、それに基づいてそれぞれの文書の特徴を計量的に表す特徴ベクトルを求めるのである。図２に、12個の文書データを分類対象とした分類事例における各文書データの特徴ベクトル算出例を示す。ベクトルの成分数は分類対象文書データ群に生起するすべての単語の種類数になるが、ここでは、単語の共生起関係を利用して３次元ベクトルに縮退させている。なお、特徴ベクトルを求めずに、単に、個々の文書毎に単語出現頻度だけを求め、文書識別符号（文書ＩＤ）に対応付けて図３に示すように記憶する構成も可能である（図３では出現頻度の記載を省略している）。分類処理部４は求められた特徴ベクトルに対してカイ自乗法、判別分析手法、またはクラスタ分析手法などを適用することにより文書分類を行う。図４に、12個の文書データをそれらの有する計量的特徴ベクトルを用いて３つの文書部分集合（クラスタ）に分類した場合の分類結果例などを示す。文書データの有する計量的な３次元ベクトルに対して例えばクラスタ分析手法の一つであるWard法などを適用することで特徴ベクトル値の近いもの同士をまとめ、３つの文書部分集合に分類することができる。つまり、各文書データは（ｂ）図に示したように３つの文書部分集合（クラスタ）のうちのいずれか一つに属する。なお、（ａ）図に示した代表値とは、所属文書データの特徴ベクトルの平均値（所属文書データの重心）である。さらに、各文書部分集合に属する文書データの各文書部分集合における順位（類似順）関係を示す特徴値として、文書データの特徴ベクトルとその文書データの属する文書部分集合の代表値との距離を求める。クラスタ（文書部分集合）12に所属している文書データ13の距離を求める例を次に示す。（$(3.00-2.66)^2+(2.00-2.00)^2+(4.00-3.66)^2)^{1/2}=0.48$図４(b)に示した距離はこのようにして求めたものである。所属している文書部分集合の代表値との距離が小さいほど、その文書部分集合に属する平均的文書との類似度が高いということになる。

【０００９】また、この実施形態の分類処理部４は複数の代表値を予め決めておき、それらの代表値との距離が小さい文書データを集めて複数の文書部分集合を求めることもできる。そのような方法では、分類処理部４は、分類対象の文書集合を構成している複数の文書の各特徴ベクトルが求まったならば、それらの特徴ベクトルの分布領域をカバーするような複数の代表値を決め、それぞれの代表値との距離が小さい文書データを集めて複数の文書部分集合を求める。また、特徴ベクトルがそのような代表値に極めて近い値になるような単語の組み合わせを求め、それぞれの組み合わせ情報を分類対象の文書集合に内在する複数の分類視点情報（内在分類視点情報）として文書解析結果保持部３に記憶しておく。あるいは、出現頻度で重みを付けられた単語の組み合わせを求め、それぞれの組み合わせ情報を内在分類視点情報とし、文書解析結果保持部３に記憶する（図５参照）。あるいは、代表値に最も近い（類似度が高い）文書中に高い頻度で出現する単語を分類視点としてもよい。また、特徴ベクトルを求めない構成では、一つの特定の単語または複数の特定の単語の組み合わせを分類視点とし、それぞれの分類視点に対応付けて文書部分集合とする（図６参照）。また、分類処理履歴保持部５には、実行した分類処理の分類視点や分類結果が保持される。利用者の分類視点指定によった分類処理を行う度毎に分類処理部４がその分類処理に識別符号（ＩＤ）を付与し、その識別符号に対応付けて指定された分類視点や分類結果情報（文書部分集合情報）を分類処理履歴保持部５に保持するのである（図７参照）。なお、図７には、分類結果情報（文書部分集合情報）として、一つの文書部分集合に分類された文書の識別符号（メンバー文書ＩＤ）を示している。また、図示の例の分類視点としては、重み付けをしていない単語を示している。未使用視点検出部８は図７に示したような分類処理履歴情報を参照することにより、これまでに用いられなかった分類視点を図５および図６に示したような内在分類視点情報中から検出する。

【００１０】図８に、分類視点を指定した文書分類時の動作フローを示す。以下、図８などに従って、この実施形態の動作を説明する。なお、分類対象の文書集合は既に文書保持部１に格納され、その文書解析が行われ、内在分類視点情報などが文書解析結果保持部３に記憶されているものとする。このような状態で、この実施形態ではまず、利用者が入力装置６および分類処理指定部７により分類視点指定の文書分類を指示する（ステップＳ

１）。そうすると、分類処理部４が未使用視点検出部８を起動して未使用視点情報を検出させる（ステップＳ２）。未使用視点検出部８は分類処理履歴保持部５に記憶されている図７に示したような分類処理履歴情報を参照することにより、これまでに用いられなかった分類視点を図５および図６に示したような内在分類視点情報中から検出（抽出）するのである。図９に、検出された未使用視点情報の一例を示す。図示していないが、代表値に最も近い文書中に高い頻度で出現する単語を分類視点とする場合には、その文書名も未使用視点情報と共に取得する。なお、対象の文書集合が文書保持部１に格納されてから初めての分類視点指定の文書分類であれば分類処理履歴情報は皆無であるので、すべての内在分類視点が未使用分類視点になる。続いて、未使用視点表示部10が、検出された未使用視点情報を表示する（ステップＳ３）。代表値に最も近い文書（代表文書）中に高い頻度で出現する単語を分類視点とする場合の表示例を図10に示す。図示したように、分類視点だけでなく、代表文書を示す情報として例えば文書名を表示させる。なお、代表文書を示す情報は文書内容の一部とかその文書のインデックス情報などであってもよい。また、分類視点を示す複数の単語は出現頻度の多い順に並べている。図10に示された各行は予め分類されたそれぞれの文書部分集合に対応しているので、利用者は、表示された複数の分類視点情報および文書名を見て、例えば所望の文書が属していると思われる文書部分集合をそのなかから探すのである。また、この実施形態では、図７に示したような分類処理履歴情報も表示させることができるので（但し、メンバー文書ＩＤは表示させない）、同様に、そのなかからも探す。その結果、未使用視点情報や分類処理結果履歴情報中に利用者の意図に合致する分類視点があればマウスなどによりそれを選択し、合致する分類視点がないと判断した場合には、ＣＰＵなどにより構成した分類視点編集手段（図示していない）が、例えば入力装置６を用いて、分類視点を構成している複数の単語の一部を削除させたり、逆に、未使用視点情報や分類処理結果履歴情報の中の他の分類視点中の単語をコピーさせて追加させたりする（ステップＳ４）。なお、そして、分類視点に修正があった場合は（ステップＳ５でYes）、分類処理部４が修正された分類視点を用いて対象の文書集合を分類し直す（ステップＳ６）。例えば、図10に示した例で、分類視点欄の「言語」と「文化」との間に「情報」という単語が追加されたならば、この文書における「情報」という単語の出現頻度を「言語」の出現頻度と「文化」の出現頻度の平均値にしてその文書の特徴ベクトルを算出し直し（つまり、修正された代表値を求める）、算出された値を既に求めてある各文書の特徴ベクトルの値と比較し、近い値の文書群を新たな文書部分集合とするのである。

【００１１】続いて、分類処理部４は、文書登録時に作成されているインデックス情報中から新たな文書部分集合に属する文書の文書名を取得し、その文書名を修正された代表値に近い特徴ベクトル値順にリストアップし、表示装置９に表示させる（ステップＳ７）。それに対して、分類視点の修正がなかった場合は（ステップＳ５でNo）、既に分類されている指定された分類視点の文書部分集合に属する文書ＩＤの文書名を取得し、その文書名をリストアップし、表示装置９に表示させる（ステップＳ７）。なお、このとき行った分類処理結果もまた分類処理履歴情報として分類処理履歴保持部５に記憶されるが、この際、分類視点に変更があった場合だけ記憶するようにすることも可能である。こうして、この実施形態によれば、広い視野から分類視点を指定することができ、したがって、分類視点の指定が一面的でなくなるので、利用者の求めている文書が表示された文書リスト中にある確率が高まり、したがって、求めている文書を容易に取得することが可能になる。なお、以上の説明において、分類視点を引き出した文書集合と前記分類視点に従って文書分類を行う文書集合とが、異なる文書集合であってもよい。例えば、先月までに文書保持部１に保持された文書集合から「問い合わせ」という単語の分類視点が未使用視点情報または分類処理結果履歴情報として引き出されたとして、今月、「新製品Ｘの機能に関する問い合わせ」という文書名の文書が前記文書集合に加わった後に、前記分類視点を用いて文書分類を行わせると、分類された文書部分集合中に「新製品Ｘの機能に関する問い合わせ」という文書も含まれるのである。以上、図１に示した文書群分類装置の場合で説明したが、本発明の文書群分類方法に従ってプログラミングしたプログラムを、例えば、着脱可能な記憶媒体に記憶させ、その記憶媒体をこれまで本発明の文書群分類を行えなかったパーソナルコンピュータなどの情報処理装置に装填することにより、その情報処理装置においても本発明の文書群分類を行うことができる。

【００１２】

【発明の効果】以上説明したように、請求項１および請求項８記載の本発明では、文書集合のそれぞれの文書データが解析されて分類処理に必要な内在情報が抽出しておかれ、分類処理に際して分類視点を指定すると、前記内在情報および指定された前記分類視点に従って文書集合が複数の文書部分集合に分類され、分類処理結果の履歴情報が保持され、保持された前記履歴情報に基づいて前記文書集合に内在するがまだ用いていない分類視点が検出されて、前記分類視点指定時に、用いられていない分類視点を表示させるようにすることができるので、表示された分類視点を参考にして分類視点の指定を行うことができ、したがって、分類視点の指定が一面的でなくなり、その結果、所望の文書部分集合を得ることができる。また、請求項２および請求項９記載の本発明では、請求項１または請求項８記載の発明において、分類処理

結果の履歴情報として、指定された分類視点情報およびその分類視点に従った分類結果である文書部分集合情報が保持されるので、前記分類視点中のいずれかを分類視点として再び指定する場合、指定された分類視点の文書部分集合情報をすばやく取り出すことができる。また、請求項３および請求項10記載の本発明では、請求項１または請求項２、または請求項８または請求項９記載の発明において、表示された用いられていない分類視点情報を用いて分類視点を指定できるので、分類視点指定作業が簡単になる。また、請求項４および請求項11記載の本発明では、請求項３または請求項10記載の発明において、さらに、分類視点情報を含んだ分類処理結果履歴情報が表示され、表示された履歴情報中の分類視点を用いて分類視点を指定できるので、さらに広い視野から分類指定を行うことができるし、所望の分類視点と同一の分類視点が表示されたなかにある確率が高くなるので、簡単に分類視点指定作業を行うことができる確率が高くなる。

【００１３】また、請求項５および請求項12記載の本発明では、請求項１乃至請求項４または請求項８乃至請求項11記載の発明において、未使用視点情報が表示される際、それぞれの未使用視点を示す一つ以上の単語および／または前記未使用視点の文書部分集合を代表する文書を示す情報が表示されるので、例えば特徴ベクトル空間を用いて文書分類を行う場合であっても、利用者は文書部分集合を示す分類視点が容易に分かる。また、請求項６および請求項13記載の本発明では、請求項１乃至請求項５または請求項８乃至請求項12記載の発明において、記憶しておいた分類視点情報を含む情報が表示され、表示された分類視点情報が編集され、編集された分類視点に従って分類処理が行われるので、表示されたなかに所望の分類視点がなくても、容易に分類視点を指定できる。また、請求項７および請求項14記載の本発明では、請求項６または請求項13記載の発明において、未使用視点情報または分類処理結果履歴情報を用いて編集し、編集された分類視点に従って分類処理が行われるので、請求項６または請求項13記載の発明の効果を実現できるだけでなく、編集のためだけに分類視点情報を含んだ特別の情報（未使用視点情報または分類処理結果履歴情報以外の情報）を表示させる必要がなくなる。また、請求項15記載の本発明では、請求項８乃至請求項14記載の発明において、分類視点を引き出した文書集合とは異なった文書集合に対して前記分類視点に従った文書分類を行うことができるので、例えば分類視点を引き出した文書集合に新たな文書が加わったりしても請求項８乃至請求項14記載の発明の効果を得ることができる。また、請求項16記載の本発明では、請求項８乃至請求項15記載の文書群分類方法に従ってプログラミングしたプログラムが例えば着脱可能な記憶媒体に記憶されるので、その記憶媒体をこれまで請求項８乃至請求項15記載の発明の文書群分類を行えなかったパーソナルコンピュータなど情報処理装置に装填することにより、その情報処理装置においても請求項８乃至請求項15記載の発明の効果を得ることができる。

【図面の簡単な説明】

【図１】本発明の一実施形態を示す文書群分類装置の構成ブロック図である。

【図２】本発明の一実施形態を示す文書群分類方法の説明図である。

【図３】本発明の一実施形態を示す文書群分類方法のデータ構成図である。

【図４】本発明の一実施形態を示す文書群分類方法の他の説明図である。

【図５】本発明の一実施形態を示す文書群分類方法の他のデータ構成図である。

【図６】本発明の一実施形態を示す文書群分類方法の他のデータ構成図である。

【図７】本発明の一実施形態を示す文書群分類方法の他のデータ構成図である。

【図８】本発明の一実施形態を示す文書群分類方法の動作フロー図である。

【図９】本発明の一実施形態を示す文書群分類方法の他の説明図である。

【図１０】本発明の一実施形態を示す文書群分類方法の画面図である。

【符号の説明】

1　文書保持部
2　文書解析部
3　文書解析結果保持部
4　分類処理部
5　分類処理履歴保持部
6　入力装置
7　分類処理指定部
8　未使用視点検出部
9　表示装置
１０　未使用視点表示部

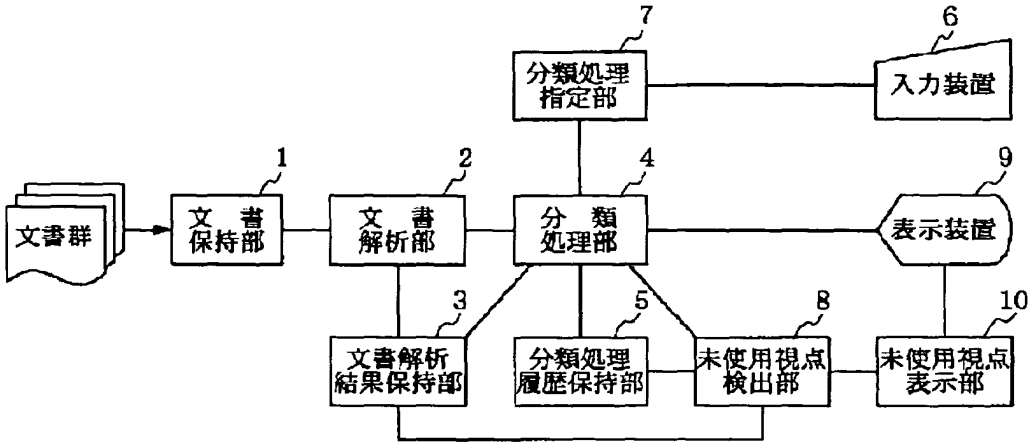
【図６】

部分集合ID	メンバー文書ID	分類視点
1	1,4,5,6,8,…	言語、ことば
2	2,3,7,…	エージェント、社会
3	2,3,5,8,…	学習

【図９】

分類視点	メンバー文書ID
囚人	2,10,…
制約	5,8,…
⋮	⋮

【図１】



【図２】

	特徴ベクトル
文書データ71	(1, 1, 1)
文書データ72	(5, 5, 5)
文書データ73	(3, 2, 4)
文書データ74	(3, 2, 3)
文書データ75	(5, 4, 6)
文書データ76	(1, 2, 1)
文書データ77	(1, 0, 1)
文書データ78	(5, 4, 5)
文書データ79	(2, 2, 4)
文書データ80	(2, 1, 1)
文書データ81	(4, 4, 6)
文書データ82	(5, 5, 6)

【図３】

文書ID	名詞句
1	知能、文化、言語、複雑系、創発、…
2	囚人、ゲーム、学習、社会的行動、エージェント、…
3	社会システム、学習、エージェント、複雑系、…
4	カオス、ニューラルネットワーク、漢字、…
5	言語、学習、制約、カテゴリー、ことば、…
6	言語、単語、複雑系、ダイナミクス、知能、創発、…
7	社会性、創発、知性、相互作用、エージェント、…
8	学習、制約、子ども、言語、曖昧さ、…
9	記憶、コンピュータ、PDA、書類、…
10	確率、メンタルモデル、教育、囚人、…
…	

【図４】

(a)

	代表値（所属文書データの重心）
クラスタ11	(1.25, 1.00, 1.00)
クラスタ12	(2.66, 2.00, 3.66)
クラスタ13	(4.80, 4.40, 5.60)

(b)

	所属クラスタ番号	距　離
文書データ71	11	0.25
文書データ72	13	0.87
文書データ73	12	0.48
文書データ74	12	0.74
文書データ75	13	0.54
文書データ76	11	1.03
文書データ77	11	1.03
文書データ78	13	0.70
文書データ79	12	0.74
文　データ80	11	0.75
文書データ81	13	0.94
文　データ82	13	0.83

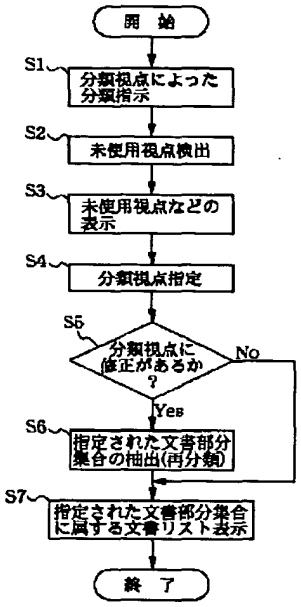
【図５】

分類視点	代表値ベクトル	組合せ
0 1	(1.00, 1.00, 1.00)	言語2, 学習2, 社会2
0 2	(1.00, 1.50, 1.50)	学習3, 社会3, 言語2
0 3	(1.00, 1.50, 2.00)	社会4, 学習3, 言語2

【図７】

分類処理ID	分類視点	分類結果（メンバー文書ID）
0 0 1	学習	2,3,5,8,…
0 0 2	社会，エージェント	2,3,7,…
0 0 3	言語，ことば	1,4,5,6,8,…

【図８】



【図１０】

分類視点	代表文書名
言語、文化、複雑系	言語と文化
脳、情報、意識	情報器官としての脳
記憶、コンピュータ、PDA	最近のコンピュータ技術
⋮	⋮